Таблица

Оценивание вероятности ошибочной классификации

Метод	До/после удаления аномальных наблюдений	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}$
Алгоритм дискриминантного анализа	Исходная выборка	14.8	46.15	25
	Выборка без аномальных наблюдений	8.00	42.8	15.6
Алгоритм логит-модели бинарного выбора	Исходная выборка	11.00	42.86	18.3
	Выборка без аномальных наблюдений	4.00	30.77	12.5

Таким образом, алгоритм классификации на основе логит-модели бинарного выбора обладает лучшими прогностическими способностями по сравнению с методом дискриминантного анализа, о чем говорят меньшие значения вероятностей ошибок первого и второго рода до и после исключения аномальных наблюдений из выборки. Аномальные наблюдения были выявлены на этапе предварительного анализа данных с помощью методов визуализации данных и алгоритмов кластерного анализа. Рассматривалась также возможность исключения аномальных наблюдений с помощью двухэтапной процедуры, основанной на использовании теста Хампеля и межклассового расстояния Махалонобиса [3]. Исключение аномальных наблюдений из выборки уменьшает вероятности ошибочных классификаций, тем самым, увеличивая количество точных классификаций заемщиков. Во всех случаях вероятность ошибки первого рода намного меньше, чем вероятность ошибки второго рода, а это значит, что вероятность пропустить «проблемного» заемщика значительно ниже вероятности отнесения «непроблемного» заемщика к классу «проблемных».

Литература

1. *Гринь Н.В., Малюгин В.И.* Исследование точности методов классификации многомерных данных в задачах кредитного скоринга // Вестник ГрГУ. Сер.2. – 2008. – №1. – С.77–85.
2. *Малюгин В.И.* Оценка устойчивости банков на основе эконометрических моделей. Банковский Вестник, 2007, Февраль.
3. *Харин Ю. С.* Робастность в статистическом распознавании образов. Минск, Университетское, 1992.

КРИПТОГРАФИЧЕСКИЕ СВОЙСТВА ГЕНЕРАТОРА МАКЛАРЕНА – МАРСАЛЬИ

И. Б. Бережной

Для быстрой и надежной защиты больших объемов данных используются поточные криптосистемы, основным элементом которых

является, как правило, генератор псевдослучайных последовательностей, от качества которого в значительной степени зависит криптостойкость алгоритма шифрования. Работа посвящена изучению некоторых криптографических свойств генератора псевдослучайных последовательностей Макларена − Марсальи [1,2].

Данный генератор состоит из таблицы T размера N и двух простейших генераторов псевдослучайных последовательностей: G_1 и G_2. Генератор G_1 порождает «заполняющую» («исходную») последовательность u над множеством $\{0,\ldots,p-1\}$, генератор G_2 – «управляющую» последовательность v над $\{0,\ldots,N-1\}$, результирующая последовательность – z над $\{0,\ldots,p-1\}$.

Если $T_j(i)$ – заполнение j-ой ячейки памяти перед началом i-го такта, то преобразование информации на i-ом такте описывается следующим образом:

$$z(i) = T_{v(i)}(i),$$

$$T_j(i+1) = \begin{cases} T_j(i), \text{если } j = v(i) \\ u(i), \text{если } j \neq v(i) \end{cases}; \quad j \in \{0,\ldots,N-1\}, \quad i = 1,2,\ldots.$$

Таким образом, последовательность v определяет адреса, по которым считываются в z и записываются в память элементы последовательности u.

Пусть i_0 – такое минимальное натуральное число, что через $i_0 - 1$ шагов в таблице T не останется начальных значений.

Лемма 1. Если v – последовательность равномерно распределенных случайных величин и

$$B(k,n) = B(k,n-1)\cdot\frac{k}{N} + B(k-1,n-1)\cdot\frac{N-k+1}{N}$$

– рекурсивная функция со следующими ограничениями:

$$B(k,k) = \frac{N!}{N^k\,(N-k)!}, \qquad B(1,n) = \frac{1}{N^{n-1}},$$

то распределение вероятностей для i_0 следующее:

$$\mathrm{P}\{i_0 = l+1\} = \frac{1}{N}B(N-1,l-1), \text{ где } l \geq N.$$

Для элементов z с индексом $i \geq i_0$ введем специальную характеристику – расстояние сдвига $L(i) = \min\{l \in N : v(i-l) = v(i)\}$. Ее смысл – расстояние между местом элемента в последовательности z и в исходной последовательности u: $z(i) = u(i - L(i))$.

Лемма 2. Если v – последовательность равномерно распределенных случайных величин, то $L(i)$ обладает следующими распределением вероятностей, математическим ожиданием и дисперсией:

$$\mathrm{P}\{L(i)=j\}=\frac{\frac{1}{N}C^{j-1}}{1-C^{i}},\quad \text{где } C=1-\frac{1}{N};$$

$$\mathrm{E}\{L(i)\}=N-\frac{iC^{i}}{1-C^{i}}\xrightarrow{i\to\infty}N;$$

$$\mathrm{D}\{L(i)\}=N^{2}-N-\frac{i^{2}C^{i}}{(1-C^{i})^{2}}\xrightarrow{i\to\infty}N^{2}-N.$$

Данные формулы показывают, что начиная с определенного места разброс значений для $L(i)$ не меняется и зависит от размера таблицы.

Теорема 1. Если в качестве G_2 применяется LFSR генератор с примитивным характеристическим многочленом, то период последовательности расстояний сдвигов определяется формулой:

$$T(\{L(i)\})=\frac{T(v)}{N-1}.$$

Лемма 3.Если последовательности u и v – периодические, то период выходной последовательности $T(z)$ удовлетворяет соотношениям:

$$\frac{HOK(T(u),T(\{L(i)\}))}{T(\{L(i)\})}\,|\,T(z)\ \text{и}\,T(z)\,|\,HOK(T(u),T(\{L(i)\}).$$

Далее под равномерной случайной последовательностью (РСП) периода τ будем понимать многократно повторенную фиксированную реализацию вектора длины τ с равномерно распределенными на $\{0,1,...,N-1\}$ элементами, которая не имеет меньшего периода .

Теорема 2. Пусть u – случайным образом (равномерно) выбранная последовательность над множеством $\{0,\ldots,p-1\}$ периода τ, v – РСП некоторого периода и последовательность сдвигов $\{L(i)\}$ имеет период t. Тогда вероятность того, что у результирующей последовательности будет период, меньший, чем $HOK(\tau,t)$, удовлетворяет следующему соотношению:

$$\mathrm{P}\{T(z)=s\cdot r\}\le\left(\frac{1}{p}+\frac{1+C^{\tau}}{(2N-1)(1-C^{\tau})}\right)^{HOK(t,\tau)-rs},$$

где $s = \dfrac{НОК(t,\tau)}{t}$, $r \mid t$, $r < t$.

*Следствие 1***.** При выполнении условий теоремы 2 $T(z) \overset{п.н.}{=} НОК(\tau, t)$.

*Следствие 2***.** В случае применения в качестве G_2 LFSR генератора с примитивным характеристическим многочленом период выходной последовательности определяетя следующей формулой:

$$T(z) \overset{п.н.}{=} НОК\left(T(u), \frac{T(v)}{N-1}\right).$$

На основании результатов серии компьютерных экспериментов была выдвинута следующая гипотеза.

Гипотеза. При использовании в качестве G_1 и G_2 LFSR генераторов на различных примитивных характеристических многочленах линейная сложность выходной последовательности $\Lambda(z)$ определяется формулой:

$$\Lambda(z) = \Lambda(u) \cdot \frac{N^{\Lambda(v)} - 1}{N - 1},$$

и для характеристических многочленов исходной и выходной последовательностей верно соотношение: $f_u(x) \mid f_z(x)$.

Также для малых значений был проведен анализ марковости выходной последовательности, который свидетельствует о высоком качестве «запутывания» исходной структуры генератора G_1.

Литература

1. *Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В*. Основы криптографии. – М.: Гелиос АРВ, 2005.
2. *Харин Ю.С., Берник В.И., Матвеев Г.В., Агиевич С.В*. Математические и компьютерные основы криптологии. – Мн.: Новое знание, 2003.

ЗАДАЧА УПРАВЛЕНИЯ РЕГИОНАЛЬНЫМИ СТРУКТУРАМИ

Н. Ю. Костюкович

В работе рассматривается трехуровневая система управления с распределенными источниками информации. Предлагается подход к автоматизации хранения, обработки и быстрой интеграции фрагментов информации в целевую предметную область [1].

Рассмотрим стандартную схему взаимодействия различного типа административных органов в рамках модели Район – Область – Республика (Рис. 1.).